ZooKeys 1249: 69–92 (2025)
DOI: 10.3897/zookeys.1249.152269

Research Article

# Comparative mitogenomics, phylogeny, and biogeography of selected species of *Saxicola* (Aves, Passeriformes)

Chuan Jiang[1,2], Wenwen Zhu[3], Shujia Wu[4], Xilong Zhao[1], Nassoro Mohamed[1,5], Bo Li[1,6]

1 *College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China*
2 *School of Ecology, Sun Yat-Sen University, Shenzhen 518107, China*
3 *College of Life Science and Technology, Jinan University, Guangzhou 510632, China*
4 *Shenzhen Frontier Science and Innovation Research Center, Shenzhen 518000, China*
5 *College of African Wildlife Management, Mweka, P.O Box 3031, Moshi, Kilimanjaro, Tanzania*
6 *State Forestry and Grassland Administration Detecting Center of Wildlife, Harbin 150040, China*

*Corresponding author: Bo Li (libo_770206@nefu.edu.cn)*

OPEN ACCESS

Academic editor:
Jorge Enrique Avendaño
Received: 7 March 2025
Accepted: 6 July 2025
Published: 13 August 2025

ZooBank: https://zoobank.org/A807F4EC-5348-4FF1-856A-D77E7C985F90

Citation: Jiang C, Zhu W, Wu S, Zhao X, Mohamed N, Li B (2025) Comparative mitogenomics, phylogeny, and biogeography of selected species of *Saxicola* (Aves, Passeriformes). ZooKeys 1249: 69–92. https://doi.org/10.3897/zookeys.1249.152269

Copyright: © Chuan Jiang et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

## Abstract

*Saxicola* is a genus of insectivorous birds widely distributed across Eurasia and Africa. Their mitochondrial genomes (mitogenomes), classification, phylogeny, and biogeography are underresearched. Here, we sequenced and/or assembled the mitogenomes of *S. stejnegeri*, *S. torquatus*, *S. maurus*, *S. rubicola hibernans*, *S. r. rubicola* and *S. dacotiae* for detailed comparative analysis, integrating them with published mitogenomes of members of Muscicapidae for phylogenetic reconstruction. Additionally, we estimated the divergence times and ancestral ranges for *Saxicola* using *ND2* and *Cytb*. These mitogenomes, spanning 16,764 bp to 16,804 bp, exhibit typical gene arrangement, AT bias, and A/C skew tendencies. Among the 13 protein-coding genes, *ND3* evolved the fastest, whereas *COX1* evolved the slowest. The 22 *tRNAs* form the classic cloverleaf structure except *trnS1* (AGY). The control region has three domains with 10 conserved elements. Mitogenomic phylogenetic analysis revealed the topology: (((((*S. r. hibernans* + *S. r. rubicola*) + *S. dacotiae*) + *S. maurus*) + *S. torquatus*) + *S. stejnegeri*), supporting the independent species status of *S. stejnegeri* rather than a subspecies of *S. maurus*. Ancestral range reconstruction indicated that *Saxicola* originated in the Indian subcontinent and mainland Southeast Asia during the late Miocene. These findings enhance our understanding of the classification, evolution, and geographical origin of *Saxicola*.

**Key words:** Classification, divergence time, mitochondrial genome, stonechat

## Introduction

*Saxicola* (stonechats) is a genus of small passerine birds belonging to the family Muscicapidae. The genus consists of widely distributed insectivorous bird species that are found across Asia, Europe, and Africa, inhabiting open habitats with woody shrubs, ranging from coastal areas to mountainous alpine environments (Winkler et al. 2020). The taxonomy of this genus remains contentious, with various checklists recognizing 13–15 species (Clements et al. 2024; Gill et al. 2024; AviList Core Team 2025), most notably the widely

distributed *S. torquatus* complex across Palaearctic and Afrotropical regions (Illera et al. 2008; Zink et al. 2009; Svensson et al. 2012; Opaev et al. 2018). Based on distribution, morphology, vocalizations, habitat preference, mitochondrial genes and genomic data (Wittmann et al. 1995; Urquhart 2002; Illera et al. 2008; Zink et al. 2009; Svensson et al. 2012; Hellström and Norevik 2014; Van Doren et al. 2017; Opaev et al. 2018; Loskot and Bakhtadze 2020; Hellström and Waern 2023), the *S. torquatus* complex is currently classified into 5–7 species: *S. stejnegeri* (or *S. maurus stejnegeri*), *S. torquatus*, *S. sibilla* (or *S. torquatus sibilla*), *S. tectes*, *S. maurus*, *S. dacotiae*, and *S. rubicola* (Clements et al. 2024; Gill et al. 2024; AviList Core Team 2025). Among them, the recognition of *S. stejnegeri* as a distinct species primarily stems from its phylogenetically distinct from the taxon *S. maurus* based on a single mitochondrial gene (Zink et al. 2009), though its relationship with *przewalskii* and *indicus* (taxa with nomenclatural priority) remains unresolved (Sangster et al. 2011; Chesser et al. 2022). Nevertheless, due to its parapatric distribution and potential hybridization with *S. maurus* (Hellström and Norevik 2014; Hellström and Waern 2023), some scholars still consider it to be a subspecies of *S. maurus* (AviList Core Team 2025). This ambiguity is also partly attributed to previous studies on this species complex relying on single mitochondrial genes, which have not yielded a conclusive phylogeny, emphasizing the need for high-resolution molecular markers to untangle this species complex.

Ancestral area reconstruction is one of the primary goals of biogeography research, which aims to investigate the geographical origins, dispersal, and evolutionary history of taxa, thereby understanding their current distribution patterns and environmental adaptations (Dupin et al. 2017). Illera et al. (2008) studied the biogeography of *Saxicola* based on *Cytb* through DIVA analysis, proposing that this genus originated in Asia, with subsequent dispersal and diversification in four main directions, eventually occupying Eurasia and Africa. However, owing to the limited phylogenetic information provided by the single *Cytb* and incomplete sampling (notably lacking samples from the eastern Palearctic), this study can serve only as a preliminary framework for the dispersal and diversification history of the genus. With continuous taxonomic revisions, the accumulation of more available gene data, and developments in statistical methods, this framework will continue to evolve and improve.

The mitochondrial (mt) genome (mitogenome), characterized by maternal inheritance in most taxa (such as birds), rapid coalescence rates, rare recombination, and suitable gene content, serves as a cornerstone for phylogenetic, taxonomic, and biogeographic studies in birds and other organisms (Vázquez-López et al. 2024). It is particularly valuable for inferring evolutionary events from the past 20 million years (Lavrov and Pett 2016). Additionally, features of mitogenomes, including nucleotide composition, codon usage, the secondary structures of RNAs, and the composition of control region (*CR*), also contain critical information about species differentiation. By comparing these characteristics among different taxa, unique and detailed insights into the evolutionary history of species can be provided.

The development of sequencing technologies has made mitogenomes available for an increasing number of avian taxa; however, this excellent marker remains poorly documented in *Saxicola*. In this study, we sequenced and/or assembled the mitogenomes of *S. stejnegeri*, *S. torquatus*, *S. maurus*, *S. dacotiae*, *S. r. hibernans*, *S. r. rubicola*, and conducted comparative analyses to characterize

their features and reveal key information on species differentiation. These data were subsequently combined with 39 mitogenomes downloaded from NCBI for phylogenetic analysis to investigate the phylogenetic relationships inter- and intra-generic *Saxicola*, providing guidance for the classification of *Saxicola* species. Finally, utilizing concatenated sequences of the *Cytb* and *ND2* genes from 14 phylogenetic species of *Saxicola*, we conducted ancestral range reconstruction to refine the origin of this genus.

## Materials and methods

### Sample collection, sequencing, and data preparation

A bird carcass, seized by the State Forestry and Grassland Administration Detecting Center of Wildlife (Harbin, China) in Beian, Heilongjiang, was identified as *S. stejnegeri* based on its location and dark plumage. Due to legal restrictions on retaining seized wildlife materials, the specimen could not be deposited in a museum, but photographs documenting key diagnostic features are publicly available (https://www.inaturalist.org/observations/292680831). Total genomic DNA was extracted from muscle tissue of this carcass via a tissue extraction kit (UElandy, Suzhou, China). The high-quality extracted DNA was first fragmented by ultrasonic shearing, and then a library with an insert size of ~350 bp was constructed on the DNBSEQ platform following the manufacturer's instructions (MGIEasy Universal DNA Library Preparation Kit, BGI). The libraries were sequenced on the DNBSEQ-T1 sequencer (MGI, China) with 150 bp paired-end strategy.

Furthermore, we downloaded the raw genome sequence fragment data for *S. maurus* (SRR17072840) (Bergeron et al. 2023), *S. r. rubicola* (ERS1563315), *S. r. hibernans* (ERS1563316), *S. torquatus* (ERS1563317) and *S. dacotiae* (ERS1563318) (Van Doren et al. 2017) from the SRA database for mitogenome assembly.

### Mitogenome assembly, annotation, and comparative analysis

All data were first filtered using fastp v. 0.23.4 (Chen 2023) with default parameters. Subsequently, the mitogenome sequences of six *Saxicola* taxa were *de novo* assembled from clean reads using GetOrganelle v. 1.7.5 (Jin et al. 2020). NOVOPlasty v. 4.3.5 (Dierckxsens et al. 2017) was also used for assembly to verify the structural correctness. Subsequently, we mapped the clean reads to the assembled mitogenome using BWA v. 0.7.18 (Jung and Han 2022) and calculated the coverage depth using Samtools v. 1.21 (Danecek et al. 2021). Mt genes were annotated using MITOS v. 2.1.9 (Zea et al. 2017). The annotation of *tRNAs* referred to the results of tRNAscan-SE v. 1.21 (Lowe and Chan 2016). Annotation of all mt genes was checked by aligning them with those of other Passeriformes species from GenBank.

The mitogenome maps were rendered in GeSeq v. 1.43 (Tillich et al. 2017). Nucleotide composition was calculated and analyzed using Geneious v. 10.1.3 (Kearse et al. 2012). Composition skew values were determined using the formulas AT-skew = [A − T]/[A + T] and GC-skew = [G − C]/[G + C] (Perna and Kocher 1995). Nonsynonymous substitution rates (Ka) and synonymous substitution rates (Ks) were calculated using DnaSP v. 6.0.7 (Rozas et al. 2017). Nucleotide diversity (*Pi*) was calculated using DnaSP v. 6.0.7 (Rozas et al. 2017) with a

window size of 100 bp and a step size of 10 bp. Relative synonymous codon usage (RSCU) and amino acid count were computed using Phylosuit v. 1.2.3 (Xiang et al. 2023). RNA secondary structures were predicted using MITOS v. 2.1.9 (Zea et al. 2017). Conserved elements in the *CRs* were analyzed with reference to previous studies (Saccone et al. 1991; Delport et al. 2002; Bi et al. 2019; Jing et al. 2020; Jiang et al. 2024). The variable sites in the *CRs* were calculated using the R package seqinr (Gouy et al. 1984).

## Mitogenomic phylogenetic analyses

To determine the phylogenetic position of the genus *Saxicola* within Muscicapidae family and enhance the robustness of the phylogeny within *Saxicola*, we retrieved the mitogenomes of 38 Muscicapidae species from GenBank (www.ncbi.nlm.nih.gov/Genbank), along with the mitogenome of a Turdidae species, *Turdus eunomus*, as an outgroup (Suppl. material 1: table S1). Notably, previous studies identified mitogenome misannotations in two Muscicapidae species on NCBI: *Cyanoptila cyanomelana* (NC_015232) is actually *Cyornis hainanus* or *C. rubeculoides*, and *Muscicapa griseisticta* (NC_045181) is *M. sibirica* (Sangster and Luksenburg 2021). Phylogenetic analysis was conducted on 13 protein-coding genes (PCGs) using both maximum likelihood (ML) and Bayesian inference (BI) methods. These PCGs were initially aligned using MAFFT v. 7.313 (Katoh and Standley 2013) and subsequently refined using MACSE v. 2.06 (Ranwez et al. 2018). The alignment sites suitable for phylogenetic analysis were selected using Gblocks v. 0.91b (Talavera and Castresana 2007) and then concatenated using the 'concatenate sequence' plugin in PhyloSuite v. 1.2.3 (Xiang et al. 2023). The optimal partitioning scheme for nucleotide substitution models was determined by PartitionFinder v. 2.0 (Lanfear et al. 2017) based on the AICc criterion (Suppl. material 2: table S2). ML analyses were performed using IQ-TREE v. 1.6.8 (Nguyen et al. 2015) with 1,000 bootstraps. BI analyses were carried out using MrBayes v. 3.2 (Ronquist et al. 2012), employing 10,000,000 generations and 4 chains, with sampling conducted every 1,000 generations. The convergence and mixing of the chains were assessed using Tracer v. 1.6.1 (Rambaut et al. 2018) to ensure that all ESS values were above 200. The consensus tree was generated after discarding the initial 25% of the trees as burn-in.

## Divergence and biogeographic analyses of *Saxicola*

To estimate the divergence time of *Saxicola* taxa, we retrieved available *ND2* and *Cytb* sequences of this genus and *Oenanthe melanoleuca* (as outgroup) from NCBI (Suppl. material 3: table S3). After obtaining the alignment and partition model using the same steps as the mitogenome phylogenetic analysis, the divergence time was estimated using BI in BEAST v. 1.8.4 (Drummond and Rambaut 2007). A Yule model was selected as the tree prior. A relaxed uncorrelated lognormal distribution was used as a clock model for all genes; clock rates (ucld.mean parameters) were estimated by implementing a prior from a published substitution rate (Weir and Schluter 2008) of 0.0105 substitutions/site/million year for *Cytb*, while a uniform prior was used for *ND2*. The secondary calibration points were selected from the crown age of *Saxicola* genus 7.6 million years ago (mya) (Zhao et al. 2023) and 4.58 mya for *S. ferreus* and *S. jer-*

*doni* from TimeTree (Kumar et al. 2022). MCMC analysis was performed twice for 50 million generations, with sampling every 1,000 generations. The convergence of each run was checked based on whether all parameters had an effective sample size greater than 200, and the first 25% of the trees were discarded as burn-in using Tracer v. 1.6.1 (Rambaut et al. 2018). The two independent runs were then combined using LogCombiner v. 2.6.4 (Bouckaert et al. 2014), and the resulting trees were summarized using TreeAnnotator v. 2.6.4 (Bouckaert et al. 2014) to generate the maximum clade credibility tree. The outgroup of the time tree was pruned using the R package ape (Paradis and Schliep 2019).

Based on the time-calibrated ultrametric tree obtained above, we used BioGeoBEARS (Matzke 2018) in R to reconstruct the most probable ancestral area for *Saxicola*. BioGeoBEARS enables the fitting of probabilistic biogeographic models using user-defined geographical areas. We defined eight geographical areas in the Old World based on the zoogeographical realms estimated in Holt et al. (2013) with adaptations: A) Indian subcontinent plus mainland Southeast Asia, B) Central Asian arid, C) Eastern Asia, D) Northeast Palearctic, E) Malay Archipelago, F) Western Palearctic, G) Afrotropical, and H) Réunion. Each stonechat taxon was coded as either present or absent in the designated geographical areas, following the breeding range information obtained from BirdLife International (Winkler et al. 2020). We tested six biogeographic models, including DEC, DIVA, and BayArea, along with their respective extensions (DEC+J, DIVALIKE + J, BayArea + J) that account for founder event speciation. The DEC, DIVA, and BayArea models all account for anagenetic processes, including dispersal and extinction, but they differ in their assumptions about cladogenetic processes (sympatry and vicariance). DEC requires that one daughter lineage is always assumed to have a range of only one area (Ree and Smith 2008). DIVA does not have this requirement but requires that daughter lineages occupy completely overlapping ranges under sympatric cladogenesis (Ronquist 1997; Yu et al. 2010). In contrast, BayArea requires that daughter lineages must inherit a range identical to their ancestor (Landis et al. 2013). The extended versions of these models incorporate a 'J' parameter, allowing one daughter lineage to inherit the ancestral range while the other disperses to occupy a new area beyond the ancestral range (Matzke 2014).

The fitting performance of the six models was evaluated through the weighted AICc (Suppl. material 4: table S4). The maximum range size was set as three since no taxa in the dataset had breeding ranges spanning more than three realms. Time-stratified analyses were employed to ensure that Réunion was not considered as the ancestral area for any species before its emergence 2.1 mya (Warren et al. 2003). Dispersal multipliers were utilized to regulate dispersal between areas, with a dispersal multiplier value of 1 for adjacent areas and 0.1 or 0.01 for nonadjacent areas (depending on the number of intervening areas) in accordance with the suggestions of Dupin et al (Dupin et al. 2017).

## Result and discussion

### Genome organization and composition

The avian mitogenome typically ranged from 16 to 17 kp, but it could be expanded due to rearrangements, pseudogenes, or tandem repeats in the *CR*, as observed in *Penelopides panini* with a length of 22,737 bp (Sammler

et al. 2011). All six mitogenomes were annotated with 37 genes each (13 PCGs, 22 *tRNAs*, and 2 *rRNAs*), with no pseudogenes or tandem repeats detected. Structural integrity was confirmed by coverage depth plots (Suppl. material 5: fig. S1), though a decrease in depth was observed in *ND6* at around 15,000 bp in each species, potentially due to sequencing biases or high heteroplasmy in that gene. Mitochondrial gene orders were conserved among species, following the *Cytb-trnT-trnP-ND6-trnE-CR* arrangement (Fig. 1a, Suppl. material 6: table S5), akin to the ancestral gene order in chicken (*Gallus gallus*) (Gibb et al. 2007), with lengths ranging from 16,764 bp to 16,804 bp, which fell within the normal range (Boore 1999). Notably, the mitogenomes of *S. r. hibernans*, *S. r. rubicola*, *S. dacotiae*, and *S. torquatus* were assembled from pooled sequencing; they represented pseudo-haplotypes of these species.

The nucleotide compositions and skewness were generally similar among the mitogenomes of six species of *Saxicola*, but differences were observed between the total genome and different gene types (Fig. 1b, Suppl. material 7: table S6). The A+T content was higher than the G+C content in both the total genome and all gene types of the six mitogenomes, indicating an A+T bias, which was most significant in *tRNAs* and the *CR*. The AT skews of the six mitogenomes were all positive (0.0777~0.0942), while the GC skews were negative (-0.3627~-0.3509), indicating strand asymmetry with more As than Ts and more Cs than Gs. For different gene types, significant A skew was observed in *rRNAs*, moderate A skew and T skew were observed in *tRNAs* and *CR*, respectively, and slight A skew (in *S. stejnegeri*, *S. torquatus* and *S. r. hibernans*) or T skew (in *S. maurus*, *S. dacotiae* and *S. r. rubicola*) was observed in PCGs. The C skew was significant in the PCGs and *CR*, moderate in *rRNAs*, and a slight G skew was observed in *tRNAs*. These features were also observed in the mitogenomes of other species of Muscicapidae (Lan et al. 2024; Yuan et al. 2024) and passerine birds in general (Yang et al. 2022; Jiang et al. 2024).

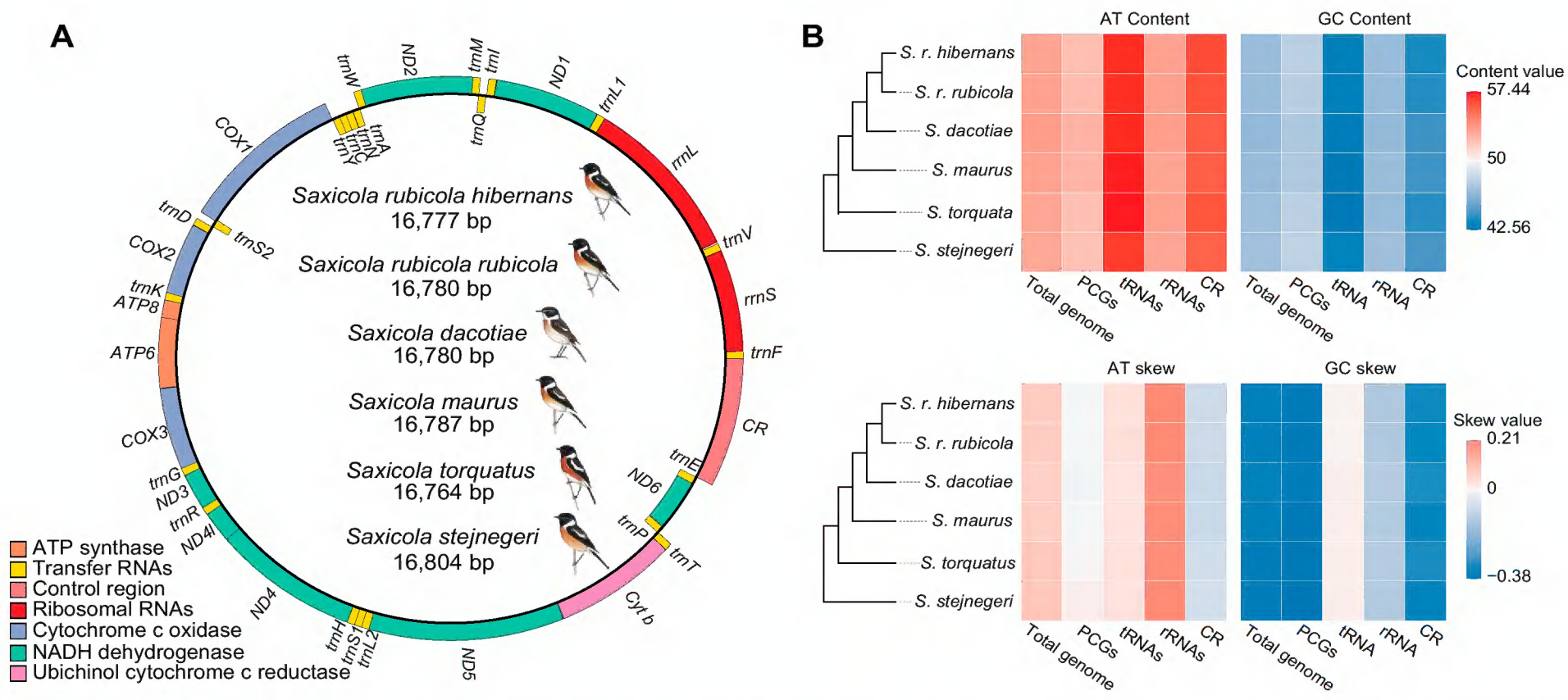


**Figure 1.** Mitogenome structure and nucleotide features. **A.** Mitogenome maps of six *Saxicola* taxa revealed a conserved ancestral gene order (Gibb et al. 2007). The genes on the H strand are outer, and the L strand genes are inner. The bird images were sourced from Birds of the World (2025); **B.** The heatmap of nucleotide features revealed interspecies similarities and differences among gene types.

## Protein-coding genes

Animal mitogenomes typically contained 13 PCGs, with *ND6* encoding on the L-strand and the remaining 12 encoding on the H-strand (Boore 1999), a pattern also observed in the mitogenomes of the six species of *Saxicola* (Fig. 1a). The lengths of these 13 PCGs were conserved across the six *Saxicola* mitogenomes, ranging from 168bp for *ATP8* to 1,818 bp for *ND5*, with a total length of 11,369 bp (Suppl. material 6: table S5).

To explore the evolutionary patterns of PCGs, we conducted analyses on nucleotide composition, skewness, selection pressure (by the Ka/Ks ratio), and *Pi* for each PCG (Fig. 2a). The nucleotide composition and skewness of PCGs were typically associated with gene function, expression levels, and DNA replication and repair mechanisms (Lercher et al. 2003; Merino et al. 2019). In our case, 13 PCGs exhibited similar AT contents, ranging from 48.61% for *ND6* to 54.76% for *ATP8*. In addition to *ND6*, which showed a G skew, the other 12 PCGs all displayed a C skew, with *ATP8* exhibiting the strongest C skew. Four PCGs (*ND1*, *ND3*, *ND6*, *COX3*) showed a T skew, most pronounced in *ND6*, while the remaining nine exhibited an A skew.

Ka and Ks represented the rates of nonsynonymous and synonymous mutations, respectively, with their Ka/Ks ratios serving as excellent indicators

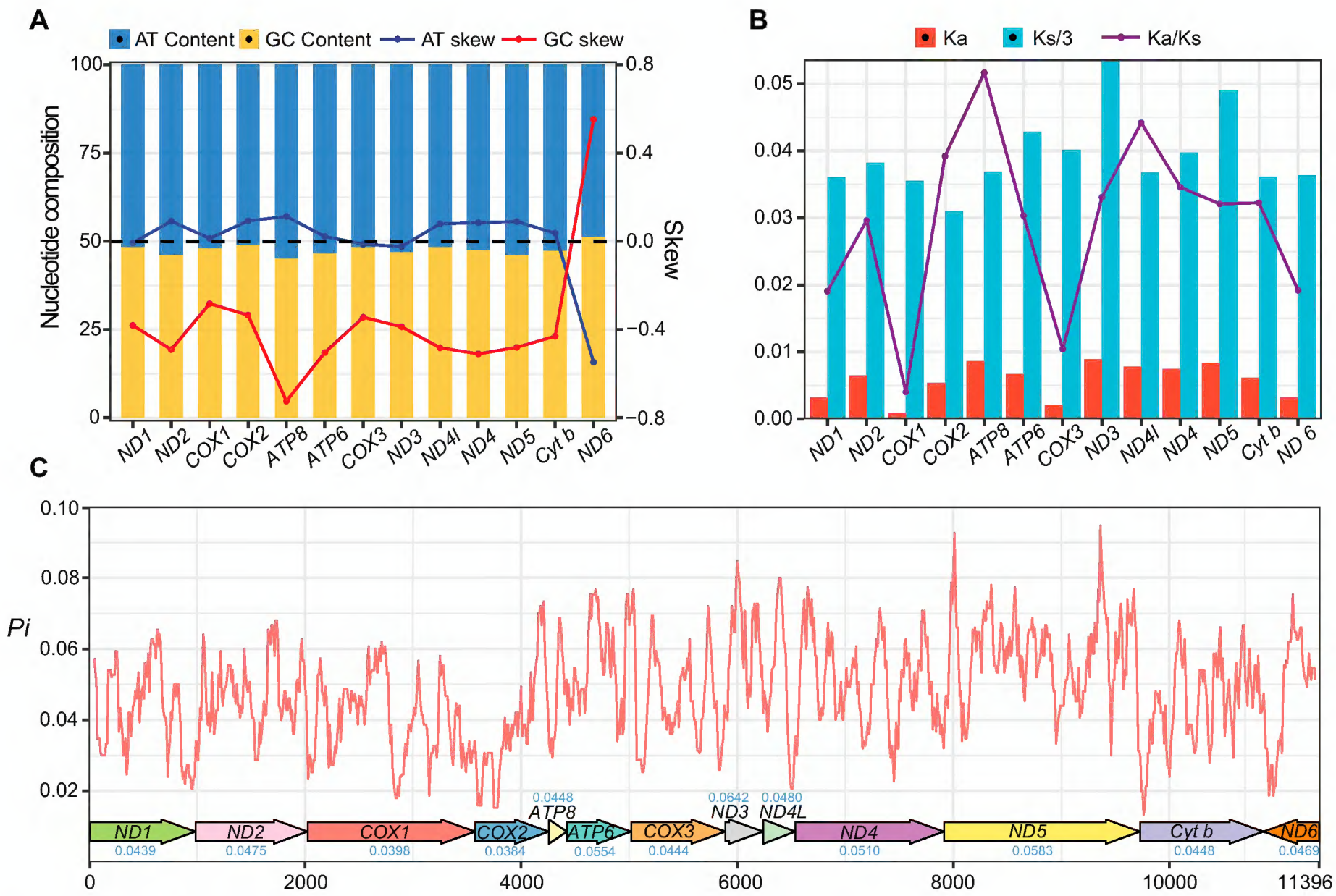


**Figure 2.** Nucleotide features, selection pressure, and nucleotide diversity of protein-coding genes (PCGs) in the six mitogenomes revealed inter-gene difference. **A.** The base composition and skewness of PCGs averaged across six mitogenomes; **B.** Ka, Ks, and Ka/Ks values of PCGs across 6 *Saxicola* taxa indicated that *ND3* evolved the fastest, while *COX1* evolved the slowest. Ka was divided by 3 (Ka/3) for better visualization; **C.** *Pi* value of the concatenated PCGs sliding window (window size = 100 bp, step size = 10 bp). The arrows below the line graph represent genes, with direction indicating gene orientation. The blue numbers above or below the arrows indicate the *Pi* for each gene.

of selection pressure on PCGs. Ka/Ks = 1 indicated neutral selection, a value greater than 1 indicated diversifying selection, and a value less than 1 indicated purifying selection (Hurst 2002). The Ka/Ks ratios of all 13 PCGs were far lower than 1 (Fig. 2b), suggesting that their evolution was under purifying selection. In particular, *COX1* had minimal Ka and Ka/Ks ratios, highlighting its high conservation and pivotal role in mitogenome functionality.

Nucleotide diversity (*Pi*) referred to the average number of nucleotide differences per site between pairs of sequences, reflecting the degree of genetic variation and selection pressure at different sites (Nei and Li 1979; Wang et al. 2024). Sliding window analysis along the concatenated sequences of the 13 PCGs revealed wide variation in *Pi* per site, with notable decreases in the latter half of *COX1* and the front half of *COX2* (Fig. 2c). For each PCG, *Pi* was lowest in *COX1* (0.0398) and *COX2* (0.0384), while highest in *ND3* (0.0642). Combining the lowest Ka/Ks ratio of *COX1* and the highest Ka value of *ND3*, it could be concluded that *COX1* was the slowest evolving gene, while *ND3* was the fastest evolving gene. The slower evolution of *COX1* further supported its utility as a barcode for identifying species of Metazoa (Bianchi and Gonçalves 2021).

We also compared the start and stop codons, RSCU, and amino acid usage of PCGs among six *Saxicola* mitogenomes. All PCGs started with ATG, while the termination codons included TAA, TAG, AGG, AGA, and T, which were conserved across species except for *ND6* (Table 1). The codon usage pattern and amino acid quantities were generally consistent among species but exhibited significant differences across different amino acids (Fig. 3), revealing the evolutionary selection pressure of different amino acids. Generally, amino acids with more synonymous codons tended to have higher usage frequencies (Bulmer 1991; Yang et al. 2024). In this study, amino acids with four synonymous codons mostly showed greater abundance compared to other amino acids, such as Leu2, Val, Ser 2, Pro, Thr, Ala, and Gly. Interestingly, despite their lower quantities, Phe and Ile showed higher abundance, suggesting potentially more stable genetic efficiency for these two amino acids (Ray et al. 2014). Furthermore, RSCU analysis indicated that codons ending with A and C were more frequent than those ending with U and G (Fig. 2a), consistent with the A skew and C skew observed in the mitogenome (Figs 1–3).

**Table 1.** Initial and terminal codons for protein-coding genes. Codons with interspecies variation are highlighted in bold.

| Species | Gene | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *ND1* | *ND2* | *COX1* | *COX2* | *ATP8* | *ATP6* | *COX3* | *ND3* | *ND4L* | *ND4* | *ND5* | *Cytb* | *ND6* |
| *Saxicola rubicola hibernans* | ATG/AGG | ATG/TAA | ATG/AGG | ATG/TAA | ATG/TAA | ATG/TAA | ATG/T | ATG/TAA | ATG/TAA | ATG/T | ATG/AGA | ATG/TAA | ATG/**TAG** |
| *Saxicola rubicola rubicola* | ATG/AGG | ATG/TAA | ATG/AGG | ATG/TAA | ATG/TAA | ATG/TAA | ATG/T | ATG/TAA | ATG/TAA | ATG/T | ATG/AGA | ATG/TAA | ATG/**TAG** |
| *Saxicola dacotiae* | ATG/AGG | ATG/TAA | ATG/AGG | ATG/TAA | ATG/TAA | ATG/TAA | ATG/T | ATG/TAA | ATG/TAA | ATG/T | ATG/AGA | ATG/TAA | ATG/**TAG** |
| *Saxicola maurus* | ATG/AGG | ATG/TAA | ATG/AGG | ATG/TAA | ATG/TAA | ATG/TAA | ATG/T | ATG/TAA | ATG/TAA | ATG/T | ATG/AGA | ATG/TAA | ATG/**TAA** |
| *Saxicola torquatus* | ATG/AGG | ATG/TAA | ATG/AGG | ATG/TAA | ATG/TAA | ATG/TAA | ATG/T | ATG/TAA | ATG/TAA | ATG/T | ATG/AGA | ATG/TAA | ATG/**TAG** |
| *Saxicola stejnegeri* | ATG/AGG | ATG/TAA | ATG/AGG | ATG/TAA | ATG/TAA | ATG/TAA | ATG/T | ATG/TAA | ATG/TAA | ATG/T | ATG/AGA | ATG/TAA | ATG/**TAG** |

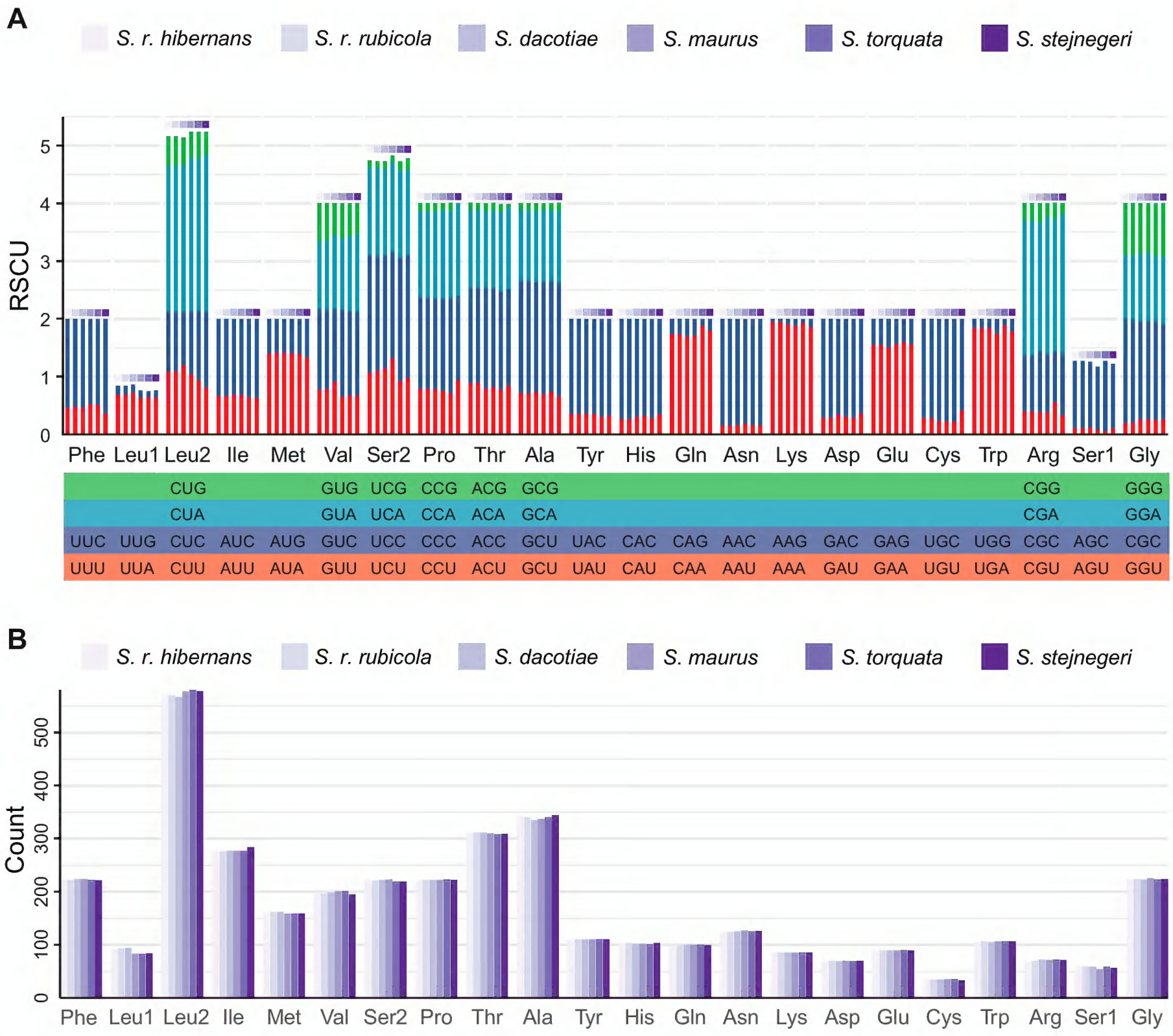
A
S. r. hibernans
S. r. rubicola
S. dacotiae
S. maurus
S. torquata
S. stejnegeri
RSCU
0
1
2
3
4
5
Phe Leu1 Leu2 Ile Met Val Ser2 Pro Thr Ala Tyr His Gln Asn Lys Asp Glu Cys Trp Arg Ser1 Gly
CUG GUG UCG CCG ACG GCG CGG GGG
CUA GUA UCA CCA ACA GCA CGA GGA
UUC UUG CUC AUC AUG GUC UCC CCC ACC GCU UAC CAC CAG AAC AAG GAC GAG UGC UGG CGC AGC CGC
UUU UUA CUU AUU AUA GUU UCU CCU ACU GCU UAU CAU CAA AAU AAA GAU GAA UGU UGA CGU AGU GGU
B
S. r. hibernans
S. r. rubicola
S. dacotiae
S. maurus
S. torquata
S. stejnegeri
Count
0
100
200
300
400
500
Phe Leu1 Leu2 Ile Met Val Ser2 Pro Thr Ala Tyr His Gln Asn Lys Asp Glu Cys Trp Arg Ser1 Gly

**Figure 3.** Codon and amino acid usage patterns in the six mitogenomes demonstrated evolutionary conservation across species. **A.** The RSCU values of the PCGs showed a preference for codons ending in A and C; **B.** Amino acid composition of the PCGs showed that amino acids with more synonymous codons were used more frequently.

## RNA genes

Similar to other avian mitogenomes (Yang et al. 2022), each of the mitogenomes of the six species of *Saxicola* contained 2 *rRNAs* and 22 *tRNAs*. Among these, eight tRNAs (*trnQ*, *trnA*, *trnN*, *trnC*, *trnY*, *trnS2*, *trnP*, and *trnE*) were encoded on the L-strand, while the rest were encoded on the H-strand (Fig. 1a).

In the six *Saxicola* mitogenomes, the length of *rrnL* ranged from 1581 to 1583, while the length of *rrnS* ranged from 982 to 984 (Suppl. material 6: table S5). Besides the varying length of *trnC*, the lengths of the *tRNAs* were conserved, ranging from 67 bp (*trnS1*) to 75 bp (*trnS2*).

To examine the differences among various RNA genes, we calculated the average nucleotide composition and skewness for each RNA gene across six *Saxicola* mitogenomes. The results revealed significant variations, with *trnC* showing the highest GC content (56.73%) and *trnQ* exhibiting the lowest (30.73%) (Fig. 4a). In terms of AT skew, 15 RNA genes presented an A skew, and nine presented a T skew. For the GC skew, 14 RNA genes presented a C skew, and ten presented a G skew. These inter-gene differences might be related to

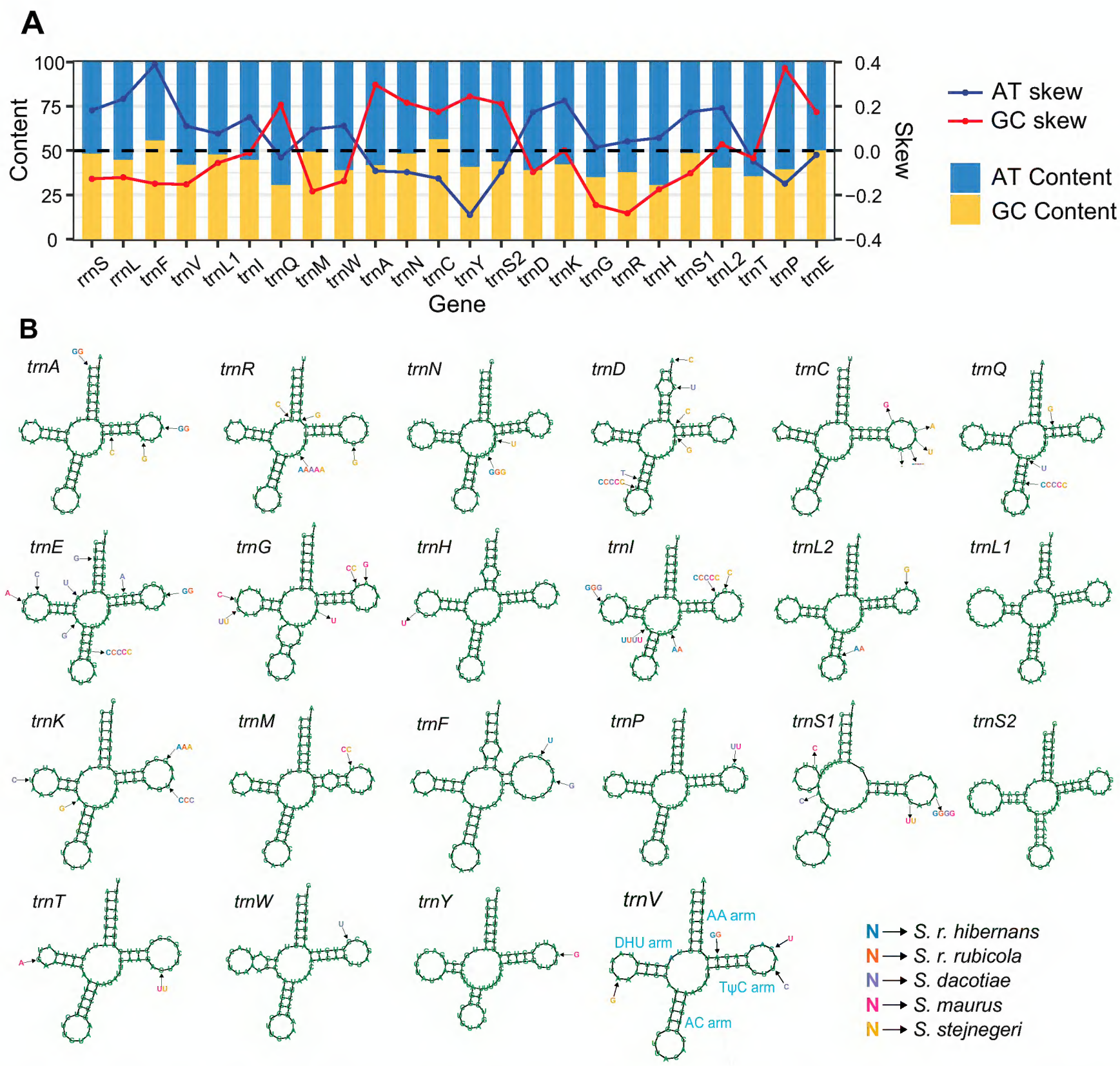


**Figure 4.** Nucleotide features of 22 *tRNAs* in six mitogenomes and their products' secondary structures. **A.** The base composition and skewness of 22 *tRNAs* averaged across six mitogenomes revealed inter-gene difference; **B.** The secondary structures of 22 *tRNAs* of *S. torquatus* revealed a conserved cloverleaf structure. The black arrows indicated variable sites between taxa, with variable bases shown in different colored letters, and '-' indicated base deletions.

the functions of different RNAs and their different locations in mitogenome (Tillier and Collins 2000; Butenko et al. 2024).

Like other birds (Yang et al. 2022), except for the absence of the dihydrouracil arm in *trnS1*, all *tRNAs* in the mitogenomes of the six species of *Saxicola* formed a cloverleaf structure consisting of the amino acid acceptor (AA) arm, anticodon (AC) arm, dihydrouracil arm (DHU arm), and TyC (T) arm (Fig. 4b). A total of 65 variable sites were detected among the mitogenomes of the six species of *Saxicola*, appearing in 20 *tRNAs*. The most frequent appearance was *trnE*, with a total of eight variable sites identified.

## Control region

The *CR* was usually considered to be the most variable in the mitogenome, and consisted of three domains: extended termination-associated sequences (ETAS), central conserved domain (CCD), and conserved sequence block (CSB)

(Saccone et al. 1991; Bi et al. 2019). This region regulated the transcription and replication of mitogenomes and contained some conserved motifs that might have function-associated secondary structures, such as CSB-1 (Delport et al. 2002). The *CRs* of the six *Saxicola* mitogenomes varied in length from 1,202 bp in *S. torquatus* to 1,244 bp in *S. stejnegeri* (Suppl. material 6: table S5). Length differences were primarily observed in the ETAS (407–409 bp) and CSB (363–405 bp), while the CCDs were conserved at 431 bp inter-specifically.

As expected, ten conserved sequence elements were identified in the *CR*, including ETAS1-2 and CSB-like in ETAS, F, E, D, C, B boxes and Bird similarity box in CCD, as well as CSB-1 and LSP/HSP in CSB (Fig. 5a). These conserved sequence boxes were also identified in other avian mitogenomes (Bi et al. 2019; Jing et al. 2020).

The evolutionary rates of the three domains varied significantly, with percentages of variable nucleotides in the three domains being 10.51%, 3.38%, and 21.99%, respectively. Most of the variable sites were concentrated in domains 1 and 3, which was consistent with previous studies (Ruokonen and Kvist 2002; Yang et al. 2021). Among the ten conserved sequence elements, CSB1-like,

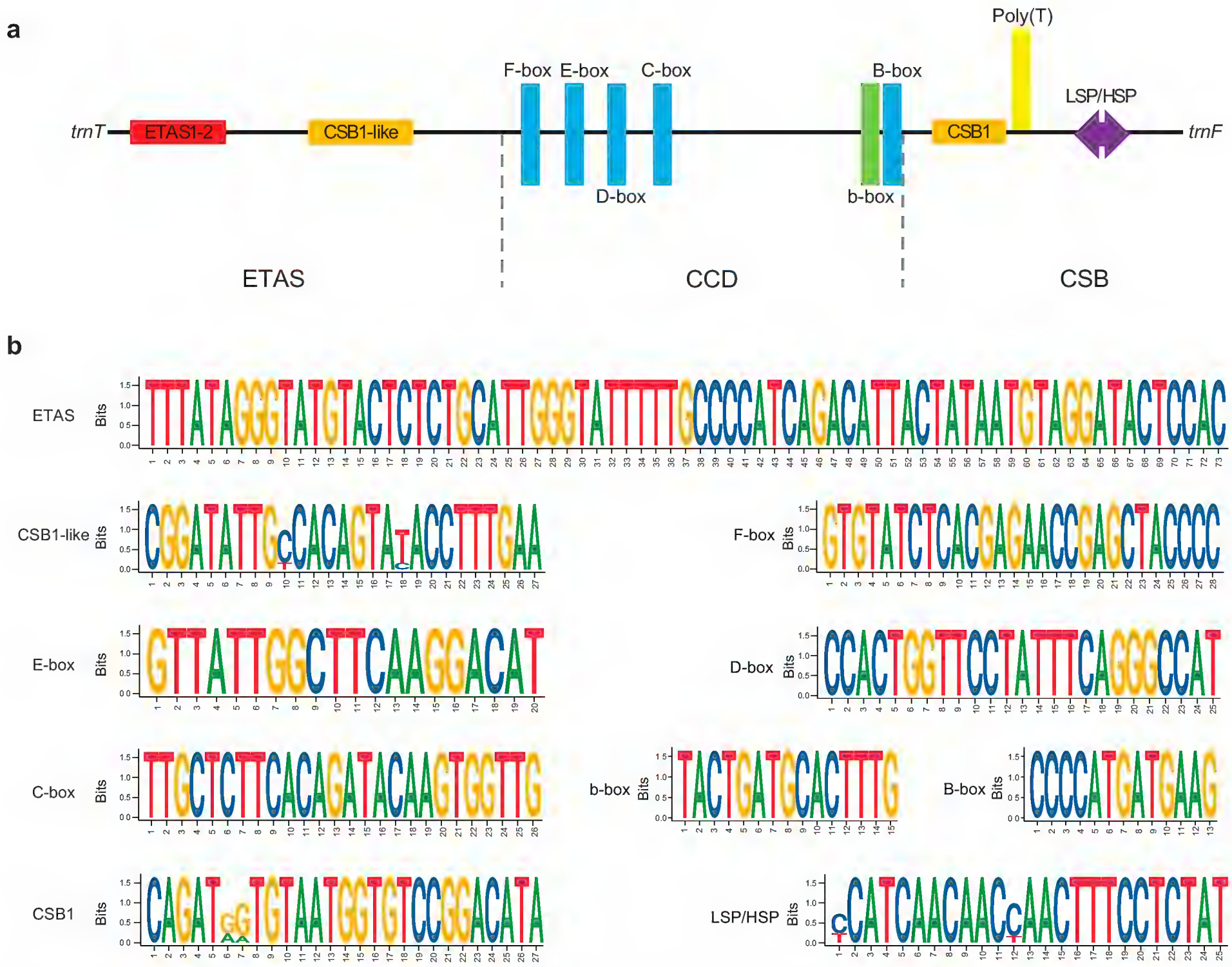


**Figure 5.** Structural features of the control region (*CR*) of six *Saxicola* mitogenomes. **A.** The structure of *CR* displayed three domains with 10 conserved elements. Different colors were used to represent distinct motifs. The size of the motifs was not scaled; **B.** The sequence logo of conserved elements in *CR* revealed variations in evolutionary rates among different elements.

CSB1, and LSP/HSP each had two variable sites, while the remaining seven were conserved across species (Fig. 5b), indicating the conserved functions of these regions (Delport et al. 2002).

## Mitogenomic phylogeny and taxonomic implications

The family Muscicapidae, comprising 303–343 species, stood as one of the largest avian families, but the phylogenetic relationships within the family were still not fully resolved (Sangster et al. 2010; Zuccon and Ericson 2010; Zhao et al. 2023). Our phylogenetic analysis included all available mitogenomes of the Muscicapidae family (as of October 2024). Except for one weakly supported node, the ML and BI approaches showed the same topology (Fig. 6, Suppl. material 8: fig. S2). As expected, all species within the same genus clustered together as monophyletic, indicating the validity of their classification. *Saxicola* was most closely related to *Oenanthe*, forming sister genera with high node support (BP=93%, PP=1), which was consistent with previous studies (Zhao et al. 2023).

Within *Saxicola*, six taxa formed a nested phylogenetic structure: (((((*S. r. hibernans* + *S. r. rubicola*) + *S. dacotiae*) + *S. maurus*) + *S. torquatus*) + *S. stejnegeri*), all with the highest node support (BP = 100%, PP = 1) (Fig. 6). Notably, *S. stejnegeri*, as the sister group of all other sampled *Saxicola* taxa, did not cluster with *S. maurus*, aligning with prior *ND2*-based findings (Zink et al. 2009). *S. maurus* and *S. stejnegeri* were phenotypically nearly identical, differing slightly in plumage color, with *S. stejnegeri* being darker (Hellström and Waern 2023), but a study found no morphological or worn spring plumage distinctions, only notable differences in male song (Opaev et al. 2018). Their distribution areas were separated along the full length of the Yenisey/Angara rivers, where extensive hybridization occurred (Vaurie 1959; Hellström and Waern 2023), blurring the boundaries between the two species. Our mitogenome-based phylogeny with the highest node support, revealed an earlier split of *S. stejnegeri* from other Palearctic, African, and island stonechats before their divergence, supporting its status as an independent species rather than a subspecies of *S. maurus*.

## Divergence and biogeography in *Saxicola*

The *Saxicola* time tree, based on concatenated sequences of *ND2* and *Cytb*, covered 14 of the 15 phylogenetic species in this genus (missing *S. macrorhynchus*) (Gill et al. 2024), and the tree topology was consistent with both the mitogenome-based results and previous research (Illera et al. 2008; Woog et al. 2008; Zink et al. 2009; Zhao et al. 2023). Using a mutation rate of 0.0105 substitutions/site/million years for *Cytb* (Weir and Schluter 2008) and secondary calibration of two nodes as priors, the BI yielded an *ND2* mutation rate of 0.0148 substitutions/site/million years, which was slightly higher than that of *Cytb* (posterior rate = 0.0109 substitutions/site/million years). Divergence dating showed that stonechats diversified from the late Miocene to the early Pleistocene, forming all species-level lineages, similar to previous estimates (Illera et al. 2008; Zhao et al. 2023).

The evaluation of the six biogeographic models indicated that adding the founder event speciation parameter J (Matzke 2014) significantly improved the model fit and the DEC+J model showed the best fit for *Saxicola* (indicated by the highest weighted AIC) (Suppl. material 4: table S4). This model's assumptions

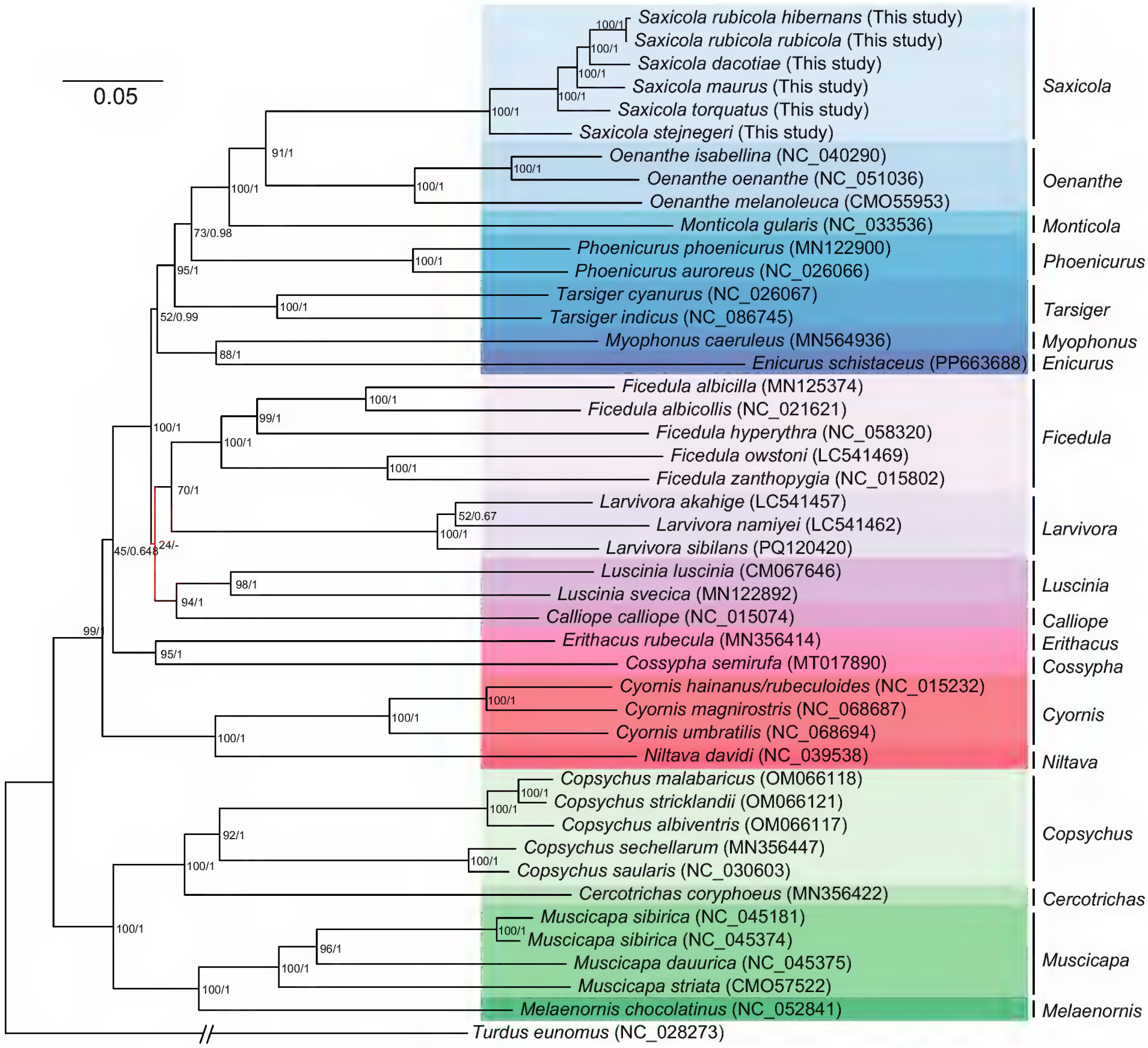


**Figure 6.** Phylogenetic tree based on 13 PCGs using ML and BI revealed the intra- and intergeneric phylogenetic relationships of *Saxicola* and supported the recognition of *S. stejnegeri* as a distinct species. The topological structure obtained from ML is displayed. Node support values are listed at nodes in the order BP/PP. Nodes highlighted in red indicate topological conflicts with the BI method (Suppl. material 8: fig. S2). Branch lengths of the outgroup are shortened for better visualization.

were the most realistic for *Saxicola* and were widely applied in biogeographic studies of other passerine birds (Oliveros et al. 2019; Batista et al. 2020; Schield et al. 2024). The biogeographic reconstruction (Fig. 7) suggested that *Saxicola* originated in the Indian subcontinent plus mainland Southeast Asia, supporting its Asian origin (Illera et al. 2008) with additional refinement. Multiple independent dispersal events and limited radiation occurred post-origin, with movements to Eastern Asia (*S. ferreus*), Western Palearctic (*S. rubetra*), Malay Archipelago (*S. gutturalis* and *S.caprata*), Central Asian arid (*S.caprata* and *S.insignis*), and Northeast Palearctic (*S. stejnegeri*). In contrast, a single dispersal to the Western Palearctic by the most recent common ancestor (MRCA) of the widely radiated clade 1 was suggested. Within clade 1, a reverse dispersal to the Indian subcontinent and an expansion into the Central Asian

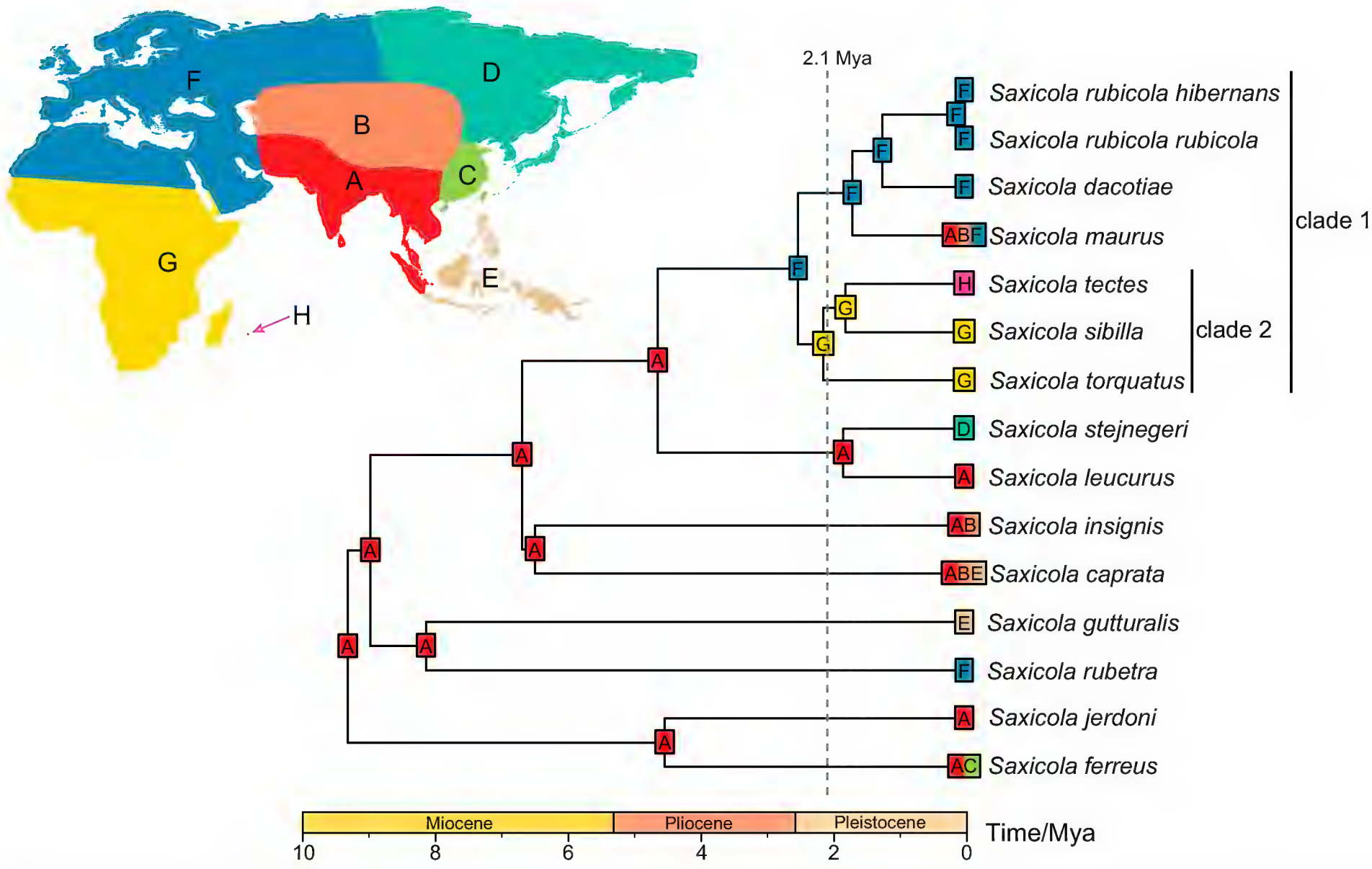


**Figure 7.** *Saxicola* biogeography analyzed with DEC + J model using *ND2 and Cytb* time-calibrated tree, supporting an origin in the Indian subcontinent and mainland Southeast Asia. The observed ranges of extant taxa are shown at the tips of the topology, with estimated ancestral ranges shown at the internal nodes. The vertical gray dashed lines indicate the timing of the uplift of Réunion Island. Letters on the map represent abbreviations for the eight biogeographic realms described in the Materials and methods section.

arid by the ancestors of *S. maurus* was proposed. This movement might have been driven by climatic fluctuations in Siberia caused by Pleistocene glacial events (Hewitt 1996, 2000), resulting in the adjacent distributions of *S. maurus* and *S. stejnegeri*, potentially facilitating hybridization between the two species (Hellström and Waern 2023). The MRCA of clade 2 dispersed from Western Palearctic to Africa, followed by a gradual colonization through Madagascar to Réunion Island. Notably, the divergence time between Madagascar stonechat *S. sibilla* and Réunion stonechat *S. tectes* was 1.86 million years ago (mya) (95% HPD = 1.28–2.47), which was close to the time (2.1 mya) of emergence of Réunion Island (Warren et al. 2003), suggesting that the uplift of the island might have facilitated the speciation of Réunion stonechat.

## Conclusions

In summary, we conducted the assembly and analysis of the mitogenomes of six *Saxicola* taxa, offering the first comprehensive insight into the mitogenomes of this genus. The six mitogenomes exhibited a conserved structure comprising 13 PCGs, 22 *tRNAs*, 2 *rRNAs*, and 1 *CR*, with characteristic AT bias and preferential A/C skewness. *ND3* was the fastest-evolving PCG, while *COX1* was the slowest. All *tRNAs* formed the classic cloverleaf structure except *trnS1* (AGY), and the *CR* had three domains with ten conserved elements. Additionally, our mitogenome-based phylogeny highlighted *S. stejnegeri* as a distinct lineage

among the six *Saxicola* taxa, affirming its species separation from *S. maurus*. Furthermore, the divergence time estimates and biogeographic reconstruction based on *ND2* and *Cytb* suggested that *Saxicola* originated in the Indian subcontinent and mainland Southeast Asia in the late Miocene, and radiated in the early Pleistocene, resulting in a wide distribution across Eurasia and Africa. Sequencing mitogenomes of other *Saxicola* lineages is needed to provide a more comprehensive characterization and to increase the accuracy of phylogenetic and biogeographic reconstruction.

## Additional information

### Conflict of interest

The authors have declared that no competing interests exist.

### Ethical statement

The bird sample was obtained from carcass related to the case, so animal ethics approval was not necessary.

### Use of AI

No use of AI was reported.

### Funding

No funding was reported.

### Author contributions

Conceptualization, C. J., B. L.; methodology, C. J., B. L.; software, C. J.; validation, C. J.; formal analysis, C. J., W. Z. and S. W.; investigation, B. L. and C. J.; resources, B. L.; data curation, C. J.; writing original draft preparation, C. J.; writing review and editing, C. J., B. L., W. Z., S. W., X. Z. and N. M.; visualization, C. J.; supervision, B. L.; project administration, B. L.; funding acquisition, B. L.. All authors have read and agreed to the published version of the manuscript.

### Author ORCIDs

Chuan Jiang https://orcid.org/0009-0002-5563-162X
Wenwen Zhu https://orcid.org/0009-0005-5433-4094
Shujia Wu https://orcid.org/0009-0006-0021-2777
Bo Li https://orcid.org/0000-0001-6824-4907

### Data availability

All data involved in this study has been uploaded to NCBI. The original sequencing data for *Saxicola stejnegeri* is under PRJNA1184580, and its annotated mitogenome is under PQ594835. The accession numbers for *Saxicola dacotiae*, *Saxicola maurus*, *Saxicola rubicola hibernans*, *Saxicola rubicola rubicola*, and *Saxicola torquatus* are BK068769, BK068770, BK068771, BK068772, and BK068773, respectively.

## References

AviList Core Team (2025) AviList: The Global Avian Checklist, v2025. https://doi.org/10.2173/avilist.v2025 [accessed 27 Jun 2025]

Batista R, Olsson U, Andermann T, Aleixo A, Ribas CC, Antonelli A (2020) Phylogenomics and biogeography of the world's thrushes (Aves, *Turdus*): New evidence for a more parsimonious evolutionary history. Proceedings. Biological Sciences 287(1919): e20192400. https://doi.org/10.1098/rspb.2019.2400

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St Leger J, Shao C, Stiller J, Gilbert MTP, Schierup MH, Zhang G (2023) Evolution of the germline mutation rate across vertebrates. Nature 615(7951): 285–291. https://doi.org/10.1038/s41586-023-05752-y

Bi D, Ding H, Wang Q, Jiang L, Lu W, Wu X, Zhu R, Zeng J, Zhou S, Yang X, Kan X (2019) Two new mitogenomes of Picidae (Aves, Piciformes): Sequence, structure and phylogenetic analyses. International Journal of Biological Macromolecules 133: 683–692. https://doi.org/10.1016/j.ijbiomac.2019.04.157

Bianchi FM, Gonçalves LT (2021) Borrowing the Pentatomomorpha tome from the DNA barcode library: Scanning the overall performance of *cox1* as a tool. Journal of Zoological Systematics and Evolutionary Research 59(5): 992–1012. https://doi.org/10.1111/jzs.12476

Birds of the World (2025) Birds of the World. In: Billerman SM, Keeney BK, Kirwan GM, Medrano F, Sly ND, Smith MG (Eds) Birds of the World. Cornell Laboratory of Ornithology, Ithaca. https://birdsoftheworld.org/bow/home

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Research 27(8): 1767–1780. https://doi.org/10.1093/nar/27.8.1767

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10(4): e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129(3): 897–907. https://doi.org/10.1093/genetics/129.3.897

Butenko A, Lukeš J, Speijer D, Wideman JG (2024) Mitochondrial genomes revisited: Why do different lineages retain different genes? BMC Biology 22(1):1– 15. https://doi.org/10.1186/s12915-024-01824-1

Chen S (2023) Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2(2): e107. https://doi.org/10.1002/imt2.107

Chesser RT, Billerman SM, Burns KJ, Cicero C, Dunn JL, Hernández-Baños BE, Jiménez RA, Kratter AW, Mason NA, Rasmussen PC, Remsen Jr JV, Stotz DF, Winker K (2022) Sixty-third supplement to the American Ornithological Society's Check-list of North American Birds. The Auk 139(3): ukac020. https://doi.org/10.1093/ornithology/ukac020

Clements JF, Rasmussen PC, Schulenberg TS, Iliff MJ, Fredericks TA, Gerbracht JA, Lepage D, Spencer A, Billerman SM, Sullivan BL, Smith M, Wood CL (2024) The eBird/Clements checklist of Birds of the World: v2024. [accessed 28 Oct 2024]

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. GigaScience 10(2): giab008. https://doi.org/10.1093/gigascience/giab008

Delport W, Ferguson JWH, Bloomer P (2002) Characterization and evolution of the mitochondrial DNA control region in Hornbills (Bucerotiformes). Journal of Molecular Evolution 54(6): 794–806. https://doi.org/10.1007/s00239-001-0083-0

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45(4): e18. https://doi.org/10.1093/nar/gkw955

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7(1): e214. https://doi.org/10.1186/1471-2148-7-214

Dupin J, Matzke NJ, Särkinen T, Knapp S, Olmstead RG, Bohs L, Smith SD (2017) Bayesian estimation of the global biogeographical history of the Solanaceae. Journal of Biogeography 44(4): 887–899. https://doi.org/10.1111/jbi.12898

Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D (2007) Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. Molecular Biology and Evolution 24(1): 269–280. https://doi.org/10.1093/molbev/msl158

Gill F, Donsker D, Rasmussen P (2024) IOC World Bird List (v14.2). https://doi.org/10.14344/IOC.ML.14.1 [accessed 28 Oct 2024]

Gouy M, Milleret F, Mugnier C, Jacobzone M, Gautier C (1984) ACNUC: A nucleic acid sequence data base and analysis system. Nucleic Acids Research 12(1Part1): 121–127. https://doi.org/10.1093/nar/12.1Part1.121

Hellström M, Norevik G (2014) The uppertail-covert pattern of 'Stejneger's Stonechat.'. British Birds 107: 692–700. https://britishbirds.co.uk/journal/article/uppertail-covert-pattern-stejnegers-stonechat [accessed 28 Oct 2024]

Hellström M, Waern M (2023) Identification of extralimital Siberian and Amur Stonechats. British Birds 116: 41–53. https://britishbirds.co.uk/journal/article/identification-extralimital-siberian-and-amur-stonechats [accessed 28 Oct 2024]

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58(3): 247–276. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405(6789): 907–913. https://doi.org/10.1038/35016000

Holt BG, Lessard J-P, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, Fabre P-H, Graham CH, Graves GR, Jønsson KA, Nogués-Bravo D, Wang Z, Whittaker RJ, Fjeldså J, Rahbek C (2013) An update of Wallace's zoogeographic regions of the world. Science 339(6115): 74–78. https://doi.org/10.1126/science.1228282

Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends in Genetics 18(9): 486–487. https://doi.org/10.1016/S0168-9525(02)02722-1

Illera JC, Richardson DS, Helm B, Atienza JC, Emerson BC (2008) Phylogenetic relationships, biogeography and speciation in the avian genus *Saxicola*. Molecular Phylogenetics and Evolution 48(3): 1145–1154. https://doi.org/10.1016/j.ympev.2008.05.016

Jiang C, Kang H, Zhou Y, Zhu W, Zhao X, Mohamed N, Li B (2024) Selected lark mitochondrial genomes provide insights into the evolution of second control region with tandem repeats in Alaudidae (Aves, Passeriformes). Life 14(7): e881. https://doi.org/10.3390/life14070881

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology 21(1): e241. https://doi.org/10.1186/s13059-020-02154-5

Jing M, Yang H, Li K, Huang L (2020) Characterization of three new mitochondrial genomes of Coraciiformes (*Megaceryle lugubris*, *Alcedo atthis*, *Halcyon smyrnensis*) and insights into their phylogenetics. Genetics and Molecular Biology 43(4): e20190392. https://doi.org/10.1590/1678-4685-gmb-2019-0392

Jung Y, Han D (2022) BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinformatics 38(9): 2404–2413. https://doi.org/10.1093/bioinformatics/btac137

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30(4): 772–780. https://doi.org/10.1093/molbev/mst010

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Gene-

ious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB (2022) TimeTree 5: An expanded resource for species divergence times. Molecular Biology and Evolution 39(8): msac174. https://doi.org/10.1093/molbev/msac174

Lan G, Yu J, Liu J, Zhang Y, Ma R, Zhou Y, Zhu B, Wei W, Liu J, Qi G (2024) Complete mitochondrial genome and phylogenetic analysis of *Tarsiger indicus* (Aves: Passeriformes: Muscicapidae). Genes 15(1): e90. https://doi.org/10.3390/genes15010090

Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP (2013) Bayesian analysis of biogeography when the number of areas is large. Systematic Biology 62(6): 789–804. https://doi.org/10.1093/sysbio/syt040

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34(3): 772–773. https://doi.org/10.1093/molbev/msw260

Lavrov DV, Pett W (2016) Animal mitochondrial DNA as we do not know it: Mt-genome organization and evolution in nonbilaterian lineages. Genome Biology and Evolution 8(9): 2896–2913. https://doi.org/10.1093/gbe/evw195

Lercher MJ, Urrutia AO, Pavlíček A, Hurst LD (2003) A unification of mosaic structures in the human genome. Human Molecular Genetics 12(19): 2411–2415. https://doi.org/10.1093/hmg/ddg251

Loskot VM, Bakhtadze GB (2020) Distribution, systematics and nomenclature of the three taxa of Common Stonechats (Aves, Passeriformes, Muscicapidae, *Saxicola*) that breed in the Caucasian region. Zoosystematica Rossica 29(1): 33–57. https://doi.org/10.31610/zsr/2020.29.1.33

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research 44(W1): W54–W57. https://doi.org/10.1093/nar/gkw413

Matzke NJ (2014) Model selection in historical biogeography reveals that founder-event speciation Is a crucial process in Island clades. Systematic Biology 63(6): 951–970. https://doi.org/10.1093/sysbio/syu056

Matzke NJ (2018) BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis with R scripts. https://doi.org/10.5281/zenodo.1478250

Merino N, Zhang S, Tomita M, Suzuki H (2019) Comparative genomics of Bacteria commonly identified in the built environment. BMC Genomics 20(1): e92. https://doi.org/10.1186/s12864-018-5389-z

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America 76(10): 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32(1): 268–274. https://doi.org/10.1093/molbev/msu300

Oliveros CH, Field DJ, Ksepka DT, Barker FK, Aleixo A, Andersen MJ, Alstrom P, Benz BW, Braun EL, Braun MJ, Bravo GA, Brumfield RT, Chesser RT, Claramunt S, Cracraft J, Cuervo AM, Derryberry EP, Glenn TC, Harvey MG, Hosner PA, Joseph L, Kimball RT, Mack AL, Miskelly CM, Peterson AT, Robbins MB, Sheldon FH, Silveira LF, Smith BT, White ND, Moyle RG, Faircloth BC (2019) Earth history and the passerine superradiation.

Proceedings of the National Academy of Sciences of the United States of America 116(16): 7916–7925. https://doi.org/10.1073/pnas.1813206116

Opaev A, Red'kin Y, Kalinin E, Golovina M (2018) Species limits in Northern Eurasian taxa of the common stonechats, *Saxicola torquatus* complex (Aves: Passeriformes, Muscicapidae). Vertebrate Zoology 68(3): 199–211. https://doi.org/10.3897/vz.68.e31606

Paradis E, Schliep K (2019) ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution 41(3): 353–358. https://doi.org/10.1007/BF01215182

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67(5): 901–904. https://doi.org/10.1093/sysbio/syy032

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F (2018) MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution 35(10): 2582–2584. https://doi.org/10.1093/molbev/msy159

Ray SK, Baruah VJ, Satapathy SS, Banerjee R (2014) Cotranslational protein folding reveals the selective use of synonymous codons along the coding sequence of a low expression gene. Journal of Genetics 93(3): 613–617. https://doi.org/10.1007/s12041-014-0429-1

Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cadogenesis. Systematic Biology 57(1): 4–14. https://doi.org/10.1080/10635150701883881

Ronquist F (1997) Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. Systematic Biology 46(1): 195–203. https://doi.org/10.1093/sysbio/46.1.195

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: Efficient bayesian phylogenetic Inference and model choice across a large model space. Systematic Biology 61(3): 539–542. https://doi.org/10.1093/sysbio/sys029

Rozas J, Ferrer-Mata A, Carlos Sanchez-DelBarrio J, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution 34(12): 3299–3302. https://doi.org/10.1093/molbev/msx248

Ruokonen M, Kvist L (2002) Structure and evolution of the avian mitochondrial control region. Molecular Phylogenetics and Evolution 23(3): 422–432. https://doi.org/10.1016/S1055-7903(02)00021-0

Saccone C, Pesole G, Sbisá E (1991) The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. Journal of Molecular Evolution 33(1): 83–91. https://doi.org/10.1007/BF02100199

Sammler S, Bleidorn C, Tiedemann R (2011) Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. BMC Genomics 12(1): 1–35. https://doi.org/10.1186/1471-2164-12-35

Sangster G, Luksenburg JA (2021) Sharp increase of problematic mitogenomes of birds: Causes, consequences, and remedies. Genome Biology and Evolution 13(9): evab210. https://doi.org/10.1093/gbe/evab210

Sangster G, Alström P, Forsmark E, Olsson U (2010) Multi-locus phylogenetic analysis of Old World chats and flycatchers reveals extensive paraphyly at family, subfamily and genus level (Aves: Muscicapidae). Molecular Phylogenetics and Evolution 57(1): 380–392. https://doi.org/10.1016/j.ympev.2010.07.008

Sangster G, Collinson JM, Crochet P-A, Knox AG, Parkin DT, Svensson L, Votier SC (2011) Taxonomic recommendations for British birds: Seventh report. The Ibis 153(4): 883–892. https://doi.org/10.1111/j.1474-919X.2011.01155.x

Schield DR, Brown CE, Shakya SB, Calabrese GM, Safran RJ, Sheldon FH (2024) Phylogeny and historical biogeography of the swallow family (Hirundinidae) inferred from comparisons of thousands of UCE loci. Molecular Phylogenetics and Evolution 197: e108111. https://doi.org/10.1016/j.ympev.2024.108111

Svensson L, Shirihai H, Frahnert S, Dickinson EC (2012) Taxonomy and nomenclature of the Stonechat complex *Saxicola torquatus sensu lato* in the Caspian region. Bulletin of the British Ornithologists' Club 132(4): 260–269. https://archive.org/details/bios-tor-246811/page/n7/mode/2up

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56(4): 564–577. https://doi.org/10.1080/10635150701472164

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45(W1): W6–W11. https://doi.org/10.1093/nar/gkx391

Tillier ERM, Collins RA (2000) The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. Journal of Molecular Evolution 50(3): 249–257. https://doi.org/10.1007/s002399910029

Urquhart E (2002) Stonechats: a guide to the genus *Saxicola*. Christopher Helm, London, 320 pp.

Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M (2017) Correlated patterns of genetic diversity and differentiation across an avian family. Molecular Ecology 26(15): 3982–3997. https://doi.org/10.1111/mec.14083

Vaurie C (1959) The birds of the Palearctic fauna: a systematic reference, order Passeriformes. https://www.semanticscholar.org/paper/The-birds-of-the-Palearctic-fauna-%3A-a-systematic-Vaurie/c003673ce2d9cb79882b6a2dc2a50c76d7797d64?sort=relevance&citationIntent=result [accessed 17 Jan 2025]

Vázquez-López M, Ramírez-Barrera SM, Terrones-Ramírez AK, Robles-Bello SM, Oca AN-M de, Ruegg K, Hernández-Baños BE (2024) Biogeographic factors contributing to the diversification of Euphoniinae (Aves, Passeriformes, Fringillidae): a phylogenetic and ancestral areas analysis. ZooKeys 1188: 169–195. https://doi.org/10.3897/zookeys.1188.107047

Wang Y, Niu S, Deng X, Bai D, Chen Z, Deng X, Huang D (2024) Genome-wide association study, population structure, and genetic diversity of the tea plant in Guizhou Plateau. BMC Plant Biology 24(1): 1–79. https://doi.org/10.1186/s12870-024-04761-x

Warren BH, Bermingham E, Bowie RCK, Prys-Jones RP, Thébaud C (2003) Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (*Nectarinia*: Nectariniidae). Molecular Phylogenetics and Evolution 29(1): 67–85. https://doi.org/10.1016/S1055-7903(03)00063-0

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Molecular Ecology 17(10): 2321–2328. https://doi.org/10.1111/j.1365-294X.2008.03742.x

Winkler DW, Billerman SM, Lovette IJ (2020) Old World Flycatchers (Muscicapidae), version 1.0. Birds of the World. https://doi.org/10.2173/bow.muscic3.01 [accessed 28 Oct 2024]

Wittmann U, Heidrich P, Wink M, Gwinner E (1995) Speciation in the stonechat (*Saxicola torquatus*) inferred from nucleotide sequences of the cytochrome-b gene. Journal of Zoological Systematics and Evolutionary Research 33(2): 116–122. https://doi.org/10.1111/j.1439-0469.1995.tb00218.x

Woog F, Wink M, Rastegar-Pouyani E, Gonzalez J, Helm B (2008) Distinct taxonomic position of the Madagascar stonechat (*Saxicola torquatus sibilla*) revealed by nucleotide sequences of mitochondrial DNA. Journal of Ornithology 149(3): 423–430. https://doi.org/10.1007/s10336-008-0290-1

Xiang C-Y, Gao F, Jakovlić I, Lei H-P, Hu Y, Zhang H, Zou H, Wang G-T, Zhang D (2023) Using PhyloSuite for molecular phylogeny and tree-based analyses. iMeta 2(1): e87. https://doi.org/10.1002/imt2.87

Yang C, Zhao L, Wang Q, Yuan H, Li X, Wang Y (2021) Mitogenome of *Alaudala cheleensis* (Passeriformes: Alaudidae) and comparative analyses of Sylvioidea mitogenomes. Zootaxa 4952(2): 331–353. https://doi.org/10.11646/zootaxa.4952.2.7

Yang C, Du X, Liu Y, Yuan H, Wang Q, Hou X, Gong H, Wang Y, Huang Y, Li X, Ye H (2022) Comparative mitogenomics of the genus *Motacilla* (Aves, Passeriformes) and its phylogenetic implications. ZooKeys 1109: 49–65. https://doi.org/10.3897/zookeys.1109.81125

Yang L, Xue J, Zhao X, Ding K, Liu Z, Wang Z, Chen J, Huang Y (2024) Mitochondrial genome characteristics reveal evolution of *Acanthopsetta nadeshnyi* (Jordan and Starks, 1904) and phylogenetic relationships. Genes 15(7): e893. https://doi.org/10.3390/genes15070893

Yu Y, Harris AJ, He X (2010) S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. Molecular Phylogenetics and Evolution 56(2): 848–850. https://doi.org/10.1016/j.ympev.2010.04.011

Yuan Z, Liu P, Lu X, Zhu D, Liu J, Guo Q, Zhang W, Duan Y (2024) Complete mitochondrial genome and phylogenetic analysis of the blue whistling thrush (*Myophonus caeruleus*). Genes 15(7): e830. https://doi.org/10.3390/genes15070830

Zea DJ, Anfossi D, Nielsen M, Marino-Buslje C (2017) MIToS.jl: Mutual information tools for protein sequence analysis in the Julia language. Bioinformatics 33(4): 564–565. https://doi.org/10.1093/bioinformatics/btw646

Zhao M, Gordon Burleigh J, Olsson U, Alström P, Kimball RT (2023) A near-complete and time-calibrated phylogeny of the Old World flycatchers, robins and chats (Aves, Muscicapidae). Molecular Phylogenetics and Evolution 178: e107646. https://doi.org/10.1016/j.ympev.2022.107646

Zink RM, Pavlova A, Drovetski S, Wink M, Rohwer S (2009) Taxonomic status and evolutionary history of the *Saxicola torquatus* complex. Molecular Phylogenetics and Evolution 52(3): 769–773. https://doi.org/10.1016/j.ympev.2009.05.016

Zuccon D, Ericson PGP (2010) A multi-gene phylogeny disentangles the chat-flycatcher complex (Aves: Muscicapidae). Zoologica Scripta 39(3): 213–224. https://doi.org/10.1111/j.1463-6409.2010.00423.x

## Supplementary material 1

### List of 45 species used for the mitogenomic phylogenetic analyses in this study

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl1

## Supplementary material 2

### The subset partitions, best models and sites used in mitogenomic phylogenetic analysis and divergence time estimation

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: docx

Explanation note: The values enclosed in parentheses following protein-coding genes indicate the positions of codons.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl2

## Supplementary material 3

### Gene list for divergence time estimation and biogeographical analysis of *Saxicola*

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: docx

Explanation note: '-' indicates that the gene is not available.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl3

## Supplementary material 4

### Support for models of geographic range evolution fitted to the time-calibrated *Saxicola* tree

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: docx

Explanation note: D = rates of range expansion, E = range contraction, J = jump dispersal weight, AICc = Akaike information criterion score.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl4

## Supplementary material 5

### Sequencing depth and coverage map of six Saxicola mitogenomes demonstrated the structural integrity

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: eps

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl5

## Supplementary material 6

### Mitogenomic structures of six *Saxicola* taxa

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: docx

Explanation note: The values separated by '/' respectively represent the characteristics of *S. r. hibernans*, *S. r. rubicola*, *S. dacotiae*, *S. maurus*, *S. torquatus*, *S. stejnegeri*.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl6

## Supplementary material 7

**Compositions and biases of the mitogenomes of six *Saxicola* taxa in various datasets of mitogenomes**

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: docx

Explanation note: The data corresponds to Fig. 1a.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl7

## Supplementary material 8

**Phylogenetic tree based on 13 PCGs using ML and BI revealed the intra- and intergeneric phylogenetic relationships of *Saxicola* and supported the recognition of *S. stejnegeri* as a distinct species**

Authors: Chuan Jiang, Wenwen Zhu, Shujia Wu, Xilong Zhao, Nassoro Mohamed, Bo Li

Data type: eps

Explanation note: The topological structure obtained from BI is displayed. Node support values are listed at nodes in the order BP/PP.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1249.152269.suppl8